

WHAT IS CLAIMED:

1. A method comprising:

specifying a model that (i) represents a set of structured data objects that include elements at particular positions, and (ii) comprises distributions of vectors, each distribution corresponding to particular positions in the respective structured data objects, each of the vectors comprising values for the particular positions, at least some distributions indicating dissimilarity at particular positions of the structured data objects; and

comparing a given set of structured data objects to the model to determine a likelihood that the given set is represented by the model.

2. The method of claim 1 in which the structured data objects comprise sequences.

3. The method of claim 1 in which the structured data objects comprise multi-dimensional maps.

4. The method of claim 3 in which the structured data objects comprise two-dimensional maps.

5. The method of claim 1 in which the structured data objects each comprises an image.

6. The method of claim 5 in which the image comprises a photographic image.

7. The method of claim 4 in which the particular positions of the two dimensional map comprise pixels.

8. The method of claim 7 in which each pixel comprises color information.

9. The method of claim 8 in which the color information comprises at least two values.

10. The method of claim 2 in which the sequences comprise audio information.

11. The method of claim 10 in which in which the audio information comprises representations of speech.

12. The method of claim 2 in which the sequences comprise financial or economic information.

13. The method of claim 2 in which the sequences comprise biopolymer sequences

14. The method of claim 2 in which elements of each of the sequences consist of a single value.

15. The method of claim 2 in which elements of each of the sequences comprise at least two values.

16. The method of claim 1 in which the model is trained to determine at least some of the distributions.

17. The method of claim 1 in which at least some others of the distributions indicate similarity.

18. The method of claim 17 in which at least some of the distributions that indicate similarity differ.

19. The method of claim 1 further comprising repeating the comparing for multiple given sets.

20. The method of claim 19 further comprising ranking the multiple given sets by the likelihoods returned by the model for each given set.

21. The method of claim 1 in which the given sets consist of two structured data objects.

22. The method of claim 19 in which the multiple given sets comprise pairwise combinations of a first and second group of structured data objects, each pairwise combination including an object of the first group and an object of the second group.

23. The method of claim 1 in which each distribution is represented as a node in a network of nodes.

24. The method of claim 23 in which the network further comprises nodes that represent an insertion or deletion in an object of the set relative to another object of the set.

25. The method of claim 23 in which interconnections between nodes are associated with a probability.

26. The method of claim 25 in which the network of nodes comprises a hidden Markov model.

27. The method of claim 23 in which the comparing comprises identifying a path that traverses the network of nodes, the path corresponding to the given set, and evaluating the likelihood of the path.

28. A medium carrying a model capable of enabling a machine to perform comparisons of a set of structured data objects to the model, the model comprising distributions of vectors, each distribution corresponding to particular positions in the respective structured data objects such that each of the vectors comprises values for the particular positions, wherein at least some distributions indicating dissimilarity at particular positions of the structured data objects.

29. An apparatus comprising a processor and storage, the storage comprising (i) a model that represents a set of structured data objects, the model comprising distributions of

vectors, each distribution corresponding to particular positions in the respective structured data objects such that each of the vectors comprises values for the particular positions, wherein at least some distributions indicating dissimilarity at particular positions of the structured data objects, and (ii) software configured to cause the processor to compare a given set of structured data objects to the model.

30. A method comprising:

specifying a model that (i) represents a set of structured data objects that include elements at particular positions, and (ii) comprises distributions of vectors, each distribution corresponding to particular positions in the respective structured data objects such that each of the vectors comprises values for the particular positions,

wherein at least some distributions indicate similarity between the structured data objects at particular positions and at least some others indicate matching to a reference structure data object at particular positions; and

comparing a given set of structured data objects to the model to determine a likelihood that the given set is represented by the model.

31. The method of claim 30 in which at least some of the distributions that indicate similarity differ.

32. A method comprising:

specifying modules, the modules comprising a first module that indicates similarity between first regions of structured data objects, and a second module that indicates dissimilarity between second regions of the objects; and

linking the modules to construct a model that indicates an ordering of regions of the similarity between the first regions and dissimilarity between the second regions.

33. The method of claim 32 in which the modules further comprise a third module that indicates matching between third regions of the objects and a reference.

34. The method of claim 33 in which the reference comprises a profile of data objects.

35. The method of claim 33 in which the reference comprises a reference data object.
36. The method of claim 32 in which at least some of the regions are discontinuous.